DOCUMENT RESUME

ED 243 901 TM 830 718

AUTHOR	Thissen, David; Wainer, Howard
TITLE	Confidence Envelopes for Item Response Theory. Technical Report No. 83-38.
INSTITUTION	Educational Testing Service, Princeton, NJ. Program Statistics Research Project.
SPONS AGENCY	Air Force Human Resources Lab., Brooks AFB, Texas.
REPORT NO	ETS-RR-83-25
PUB DATE	May 83
CONTRACT	F41689-82-C-0020
NOTE	29p.
AVAILABLE FROM	Educational Testing Service, Research Publications R-116, Princeton, NJ 08541.
PUB TYPE	Reports - Research/Technical (143)
EDRS PRICE	MF01/PC02 Plus Postage.
DESCRIPTORS	*Graphs; *Latent Trait Theory; *Mathematical Formulas; Mathematical Models; Regression (Statistics); Sampling
IDENTIFIERS	*Confidence Intervals (Statistics)

ABSTRACT

A statistical method is described and illustrated which provides confidence envelopes around item response functions. Examples of 95 percent confidence envelopes for the one-, two-, and three-parameter logistic response models are given. In addition, the authors describe N-line plots, which show the genesis of the envelope as well as the density of lines in the confidence region. These too are illustrated for the one-, two-, and three-parameter logistic models. (Author)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

Confidence Envelopes for Item Response Theory

David Thissen

and

Howard Wainer

Program Statistics Research
Technical Report No. 83-38

Research Report RR-83-25

May 1983

Not to be cited or further reproduced
without the permission of the authors.

Copyright © 1983 by Educational Testing Service. All rights reserved.

Abstract

A statistical method is described and illustrated which provides confidence envelopes around item response functions. Examples of 95% confidence envelopes for the one-, two-, and three-parameter logistic response models are given.

In addition, we describe, N-line plots, which show the genesis of the envelope as well as the density of lines in the confidence region. These too are illustrated for the one-, two-, and three-parameter logistic models.

I. Introduction

In seminal work published in 1929, Working and Hotelling developed the formulas for the extrapolation of the concept of the confidence interval to regression lines. They provided a "confidence envelope" enclosing the population line with probability $1-\alpha$. Such an envelope for a regression function $F(\cdot)$ has also been called a confidence "band" by Miller (1966). It is defined by the functions $L(\cdot)$ and $U(\cdot)$ illustrated in Figure 1 satisfying the relationship

$$P\{U(\cdot)>F(\cdot)>L(\cdot),\ \forall\ x\} = 1-\alpha \qquad (1)$$

in which $U(\cdot)$ and $L(\cdot)$ are upper and lower bounds of the envelope

- - - - - - - - - - -

Figure 1 about here

- - - - - - - - - - -

respectively, and $F(\cdot)$ is the function. The envelope problem for linear models has been treated extensively by Working and Hotelling (1929), Roy (1957), Miller (1966), and others; but all use the intimate relationships of linear models with multinormal error to develop computational formulae for $U(\cdot)$ and $L(\cdot)$. Confidence banks for logistic regression have been developed by Brand, Pinnock, & Jackson (1975) and Hauck (1983), and for more general CDF's by Cheng & Iles (1983), but none of these developments include the three-parameter logistic commonly used in test theory. In this paper we describe and illustrate a method for constructing confidence envelopes satisfying (1) for any regression function $F(\cdot)$ which is monotonic in its parameters.

It is easier to understand the nonlinear problem if we first consider the

basic concepts underlying the standard linear regression envelope. This requires the simultaneous consideration of two spaces in a duality diagram: the "function space" on the right of Figure 2 and the "parameter space" on the left. While there are a number of algebraic procedures for the derivation of

- - - - - - - - - - -

Figure 2 about here

- - - - - - - - - - -

the hyperbolic envelope boundaries shown in Figure 2 (Miller, 1966) we will consider a geometric interpretation here.

The left panel of Figure 2 illustrates the conventional sampling space for the parameters of linear regression. The circle[1] encloses the central 95% of a bivariate normal with its mean equal to the estimates and covariance matrix given by classical maximum likelihood results. The phrase "95% confidence interval" describes the probability that the circle "covers" the true value of the parameter vector, so the circle in the left panel of Figure 2 illustrates a confidence region for the slope-intercept combination.

However, this illustration of the confidence region in the parameter space is not very informative. It is often more useful to map the confidence region into the function space. Each point of the parameter space corresponds to a stet in the function space. The center of the parameter space corresponds to the estimated regression line, and all lines represented by slope-intercept combinations within the circle are in the envelope described by the hyperbolae in Figure 2. Thus, the hyperbolic boundaries describe a 95% envelope for the regression line because they bound lines corresponding to the central 95% confidence retion in the parameter space.

Indeed, each point on the hyperbolae corresponds to a point (a set of parameters) on the boundary of the circle in the parameter space. For instance, the point marked 1 on the lower hyperbola represents a point with low slope and low intercept — marked 1 in the parameter space — so that it gives the lowest value of $a+bx_1$ for an a and b in the central 95% region of the parameter space. So the 95% envelope pictured in Figure 2 is simply a representation in the function space of the circle in the parameter space. One could plot the lower hyperbola in Figure 2 by plotting, for each value x_0, the values $y_0 = a_{x_0} + b_{x_0} x_0$ where (a_{x_0}, b_{x_0}) is the point on the circle which minimizes y. There is a closed form solution for the linear case, so numerical minimization is not required. But, the principle generalizes readily to nonlinear regression; explicit minimization is required with some models to find the envelope boundary.

II. Logistic Item Response Models

We (Thissen and Wainer, 1982b) have previously described a computational technique for plotting the boundaries for the regression equivalents of two- and three-parameter logistic functions. This is similar to the method independently developed by Cheng and Iles (1983) for confidence bands for cumulative distribution functions, and is equivalent to Brand, Pinnock and Jackson (1973) for the two-parameter logistic. The problem is described schematically in Figure 3. Any item response model may be denoted $F(\theta;\underline{v})$ in

- - - - - - - - - - - - - - -

Insert Figure 3 about here

- - - - - - - - - - - - - - -

which θ is the latent ability variable and $\underline{v}$ represents the parameters of the nonlinear regression model; in the more usual notation, $\underline{v}=[a,b]$ for the two-parameter logistic model and $\underline{v}=[a,b,c]$ for the three-parameter model. The central 95% confidence region of the sampling distribution for the estimates of the parameter vector $\underline{v}$ is illustrated by the bivariate normal ellipse in the top panel of Figure 3. Error covariance between the parameters is suggested in the ellipse because the errors of estimate of the parameters of item response functions are correlated in most circumstances; the single exception is when $E(\theta)=b$ (Thissen and Wainer, 1982a). The non-linear envelope problem may be stated in the same way as in the linear case: to plot the lower boundary $L(\theta;\hat{\underline{v}})$ for parameter estimates $\hat{\underline{v}}$, find for each value θ_0 the minimum value of $F(\theta_0;\underline{v}^*)$ where $\underline{v}^*$ is constrained to lie on the boundary of the parameter space.

There are a number of ways to solve such a constrained minimization problem. Our solution (Thissen and Wainer, 1982b) was to reparameterize the problem to become a minimization problem in terms of the angle defining each vector $\underline{v}^*$. Once the minimization problem is solved, the resulting points are plotted as $L(\theta;\hat{\underline{v}})$. The negative of the same function is minimized to give $U(\theta;\hat{\underline{v}})$.

Figure 4 is an illustration of the resulting 95%-envelope for a two-parameter logistic of the form

$$F(\theta;[a,b]) = 1 / \{1+\exp[-a(\theta-b)]\}$$

with parameters [1,0] and a sample of size 100 drawn from a standard normal distribution. It graphically suggests the wide variety of ICCs which might fit the data under these circumstances. Figure 5 shows the much narrower bounds given by a sample size of 500, and a steeper slope. The envelope in Figure 5 is not symmetrical, because the location is -1 and that is not the mean of the population distribution, which is zero. Thissen and Wainer (1982a) show that if the location of the item is not the same as the mean of the population distribution, the sampling errors of a and b are correlated, and that covariance translates into assymmetry in the regression envelope. By contrast, the envelope for Figure 4 is symmetrical since b=0 there.

- - - - - - - - - - - - - - - - - - - -

Insert figures 4, 5 and 6 about here

- - - - - - - - - - - - - - - - - - - -

Figure 6 illustrates the envelope for the three-parameter logistic,

$$F(\theta;[a,b,c]) = c + (1-c) / \{1+\exp[-a(\theta-b)]\}.$$

While such plots can be made, the process is complicated by the existence of local minima in the minimization problem described earlier. Lord and Stocking (personal communication) have shown that the envelope problem has local minima for the three-parameter logistic model, so great care is required in the computation to assure finding the true minimum to plot. Nevertheless, such plots are potentially useful tools. Note that the envelopes constructed will be too narrow (i.e., the estimates will look optimistically good) if the solution is a local rather than a global minimum.

III. Density of the curves within the envelope: N-line plots

A Bayesian interpretation would denote the sampling distribution of the parameters in a regression problem as the posterior density of the parameters obtained from a suitable uninformative prior distribution. In the Bayesian view, parameter vectors near the estimates are more likely (have greater posterior density) than parameters more distant from the posterior mean, although all may lie within the central 95% region. While plots of the $(1-\alpha)$-envelope define precisely a $(1-\alpha)$ -confidence interval for the fitted curve, they do not graphically display the density of the curves within the envelope. Indeed, for a frequentist statistician there is no "density of curves within the envelope"; but the Bayesian formulation derives such a concept from the posterior density. N-line plots are plots which (approximately) fill the $(1-\alpha)$-envelope with curves, in proportions roughly corresponding to their posterior likelihood. N-line plots consist of plots of N lines drawn using parameters randomly sampled from the posterior distribution for the parameters. N-line plots are constructed in such a way that, at any point on the abscissa, the most extreme of the several lines is located in expectation, on the boundary of the 95% envelope. In practice, of course, the extreme lines or N-line plots wander slightly inside and outside the actual boundary.

III.1 A One-Parameter Example: 25-line Plots

The simplest widely used item response model is the one-parameter logistic model, in which (for many practical purposes) the slope may be taken

to be a known constant; this is similar to a regression model of the form

$$F(\theta;b) = 1 \,/\, \{1+\exp[-(\theta-b)]\}$$

in which b is a location parameter, and the only estimated parameter in the model. The computation of the $(1-\alpha)$ - envelope for the fitted curve is trivial: if $\hat{b}$ is the estimate of b, with a standard error of σ_b, then the upper boundary for the 95% envelope $U(\cdot)$ is $F(\theta;b-1.96\sigma_b)$; and the lower boundary for the envelope is $L(\cdot) = F(\theta;\hat{b}-1.96\sigma_b)$. A typical 95% envelope for such a function is plotted in Figure 7, with the modal (fitted) curve in the center of the envelope. The boundaries of the envelope plot describe the central 95% confidence interval for the fitted curve.

- - - - - - - - - - - - - - - - - -

Insert Figures 7 and 8 about here

- - - - - - - - - - - - - - - - - -

In Figure 8, following our description of the posterior density for b as $N(\hat{b},\sigma_b^2)$, we have plotted the curves corresponding to 25 random deviates from a normal density with mean $\hat{b}$ and standard deviation σ_b. Note that such randomly sampled curves are dense near the curve corresponding to $\hat{b}$ and, albeit more thinly toward the edges, fill the 95% envelope of Figure 7.

The envelope described by the upper and lower extreme curves of the 25-line plot is, in expectation, almost exactly equal to the true 95% envelope. This is because the distance between the upper and lower boundaries of the 95% envelope plot is determined by plotting $F(\theta;\hat{b}+1.96\sigma_b)$: plotting F for two values of b, centered on $\hat{b}$, which have a range of $2(1.96)\sigma_b$. The

random deviates which yield the 25-line plot are also centered on $\hat{b}$, and the expected value of their range is $3.93\sigma_b$, which is very nearly $2(1.96)\sigma_b$. (Values of the expected value of the range of various size samples from normal distributions are given by Pearson and Hartley [1956]). Thus the boundaries of the 95% envelope and expected values of the highest and lowest curves of the 25-line plote are in the same place. The 25-line plot gives a clearer graphic description of the envelope for the curve.

III.2 A Two-Parameter Example: 85-line Plots

The regression version of the two-parameter logistic model, in which

$$F(\theta;a,b) = 1 \,/\, \{1+\exp[-a(\theta-b)]\}$$

provides more interesting N-line plots, but they are more costly to make. The vector [a,b] is assumed estimated with posterior density $N([\hat{a},\hat{b}],\Sigma_{ab})$. The upper and lower extreme curves of the N-line plot are equal in expectation, to the boundaries of the 95% envelope. The appropriate N-line plot has, as its (squared) range on any line through the origin of the density (shown in Figure 3) Mahalanobis distance $\chi^2_{(1-\alpha)}$ (2). Thus, on any line through the origin of the two-dimensional space in Figure 3, the expected range of the projections of the random deviates of the N-line plots are the same as the distance between the two boundaries of the central confidence region on that line. Every value of θ has as its upper and lower $(1-\alpha)$-envelope boundaries the parameters at the intersections of the $(1-\alpha)$ confidence region boundary with some line through the origin of the [a,b]-space. So the requirement that the expected value of the range of the projections of the random sample on any such line be equal to the $(1-\alpha)$ confidence region boundary difference

given the desired result: the expected location of the upper and lower N-line curves are along the upper and lower $(1-\alpha)$-envelope boundaries.

For the two-dimensional case, N (regrettally) is 85. This is determined by noting that the [a,b] space of Figure 3 is a projection of the standard two-dimensional multinormal, and the central 95% region of the standard bivariate normal has radius equal to the square root of the value of a chi-square on two degrees of freedom associated with 95% (i.e $\sqrt{\chi^2_{(.95)}(2)}$, which equals 2.45. The expected value of the standard normal deviates sampled must be twice this (4.9) to provide an N-line plot of the required breadth. Pearson and Hartley's (1956) table indicates that the required sample size is is 85. Consequently we can make N-line plots from the two-space by sampling 85 standard normal deviates, transforming them to have the [a,b] density in Figure 3, and plotting the resulting curves. Figure 9 shows an illustrative 85-line plot; the corresponding 95%-envelope plot was in Figure 4. The 85-line plot shows clearly how the odd shape of the 95%-envelope plot arises from the distribution of curves. Flat curves make up U(•) on the left and L(•) on the right, while steep curves make up L(•) on the left and U(•) on the right.

- - - - - - - - - - - - - - - - - - -

Insert Figures 9 and 10 about here

- - - - - - - - - - - - - - - - - - -

Eighty-five lines are a lot to plot. Twenty-five line plots are equivalent (in expectation) to only 85% envelopes but, as Figure 10 shows, they give much of the detail available in 85-line plots for a third of the cost.

III.3 The Three-Parameter Model

Through similar arguments to those of the preceding sections, N-line plots for t' ree-parameter models require N=235 to match (in expectation) 95%-envelope plots. Figure 11 shows a 235-line plot for the three parameter logistic model described above, and Figure 12 is the 95%-envelope plot. Two

- - - - - - - - - - - - - - - - - - -

Insert figures 11 and 12 about here

- - - - - - - - - - - - - - - - - - -

hundred thirty-five lines take a plotter a long time. The width of a 25-line plot for three-parameter models is equivalent to a 72%-envelope, so the much simpler plots could be useful in many applicatons.

IV. Conclusions

A description of the sampling variation of a fitted function in terms of the standard errors or the variance of the posterior density of the parameters may not be the most useful specification of that variation. Functions are usually fitted for the sake of the functions, not the parameters. This is especially true of item response functions when they are to be concatenated to form posterior density functions for ability associated with certain observed response vectors.

Confidence envelopes provide a description of the sampling variation of item response curves in the space of the fitted function. They can be used to give the data analyst a clear idea of the class of item response curves which are compatible with the data. N-line plots may be used to show the width of the envelope, as well as the slopes and relative posterior density of the included curves. These techniques can be quite useful in applied item analysis.

Acknowledgement

This work was supported, in part, by the Air Force Human Resources Laboratory in contracts F41689-81-6-0012 to McFann-Gray Associates and F41689-82-C-0020 to the Educational Testing Service. We were helped toward the solution of this problem and toward clearer prose in its description through conversations and comments on earlier drafts by R. Darrell Bock, Henry Braun, Neil Dorans, Benjamin Fairbank, Paul Holland, Neal Kingston, Frederic Lord, Donald Rubin, and Glenn Shafer. Fuzziness that remains, in thought or in prose, is, of course, our own.

Footnotes

[1] In general this is an ellipse with major and minor axes of length equal to $2.45\sigma_a$ and $2.45\sigma_b$. The coefficient 2.45 equals $\sqrt{\chi^2(.95)}$ with 2df; see Thissen and Wainer, 1982b for a more complete explanation. For purposes of exposition, we are using a rescaling that yields a circle. This can be done without loss of generality.

References

Brand, R.J., Pinnock, D.E., & Jackson, K.L. Large sample confidence bands for the logistic response curve and its inverse. American Statistician, 1973, 27, 157-160.

Cheng, R.C.H. & Lles, J.C. Confidence bands for cumulative distribution functions and continuous random variables, Techmometrics, 1983, 25, 77-86.

Hauck, W.W. A note on confidence bands for the logistic response curve. American Statistician, 1983, 37, 158-160.

Miller, Jr., R.G. Simultaneous Statistical Inference. New York: McGraw-Hill, 1966.

Pearson, E.S., & Hartley, H.O. Biometrika Tables for Statisticians, Vol. 1, 3rd Edition. London: Cambridge University Press, 1966.

Roy, S.N. Some Aspects of Multivariate Analysis. New York: John Wiley & Sons, 1957.

Thissen, D., & Wainer, H. Some standard errors in item response theory. Psychometrika 1982a, 47, 397-412.

Thissen, D., & Wainer, H. Confidence envelopes for monotonic functions: Principles, derivations, and examples. Air Force Human Resources Laboratory Technical Paper No. 82-37, 1982b.

Working, H. & Hotelling, H. Application of the theory of error to the interpretation of trends. Journal of the American Statistical Association, Suppl., 1929, 24, 73-85.

Figure Captions

Figure 1. Generalized illustration of a confidence envelope for a fitted curve $\hat{F}(\cdot)$, in which the envelope is bounded above by $U(\cdot)$ and below by $L(\cdot)$.

Figure 2. The confidence envelope in linear regression: the left panel shows the central 95% region in the parameter (slope, intercept) space, while the right panel shows the corresponding hyperbolic envelope for the fitted line in the function space.

Figure 3. The confidence envelope generalized to nonlinear regression: the upper panel shows the central 95% region in the parameter space (for parameters with correlated error) and the lower panel shows the corresponding envelope.

Figure 4. A 95% confidence envelope for a two-parameter logistic ICC, with a=1, b=0, and sample size 100.

Figure 5. A 95% confidence envelope for a two-parameter logistic ICC, with a=1.5, b=-1, and sample size 500.

Figure 6. A 95% confidence envelope for a three-parameter logistic ICC, with a=1.5, b=-1, c=0.1, and sample size 1000.

Figure 7. A 95% confidence envelope for a one-parameter logistic ICC, with b=0 and sample size 100.

Figure 8. Twenty-five one-parameter logistic ICCs, with locations sampled randomly from the posterior distribution for b=0 and sample size 100.

Figure 9. Eighty-five two-parameter logistic ICCs, with parameters sampled from the posterior distribution for a=1, b=0, and sample size 100.

Figure 10. Twenty-five two-parameter logistic ICCs, with parameters sampled from the posterior distribution for a=1, b=0, and sample size 100.

Figure 11. Two hundred thirty-five three-parameter logistic ICCs, with parameters sampled from the posterior distribution for a=1.5, b=1, c=0.1, and sample size 1000.

Figure 12. A 95% confidence envelope for the three-parameter ICC of Figure 11.

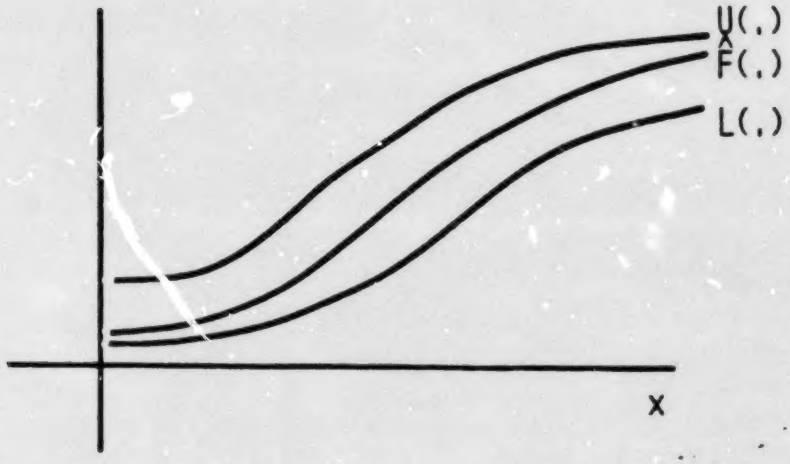


FIGURE 1

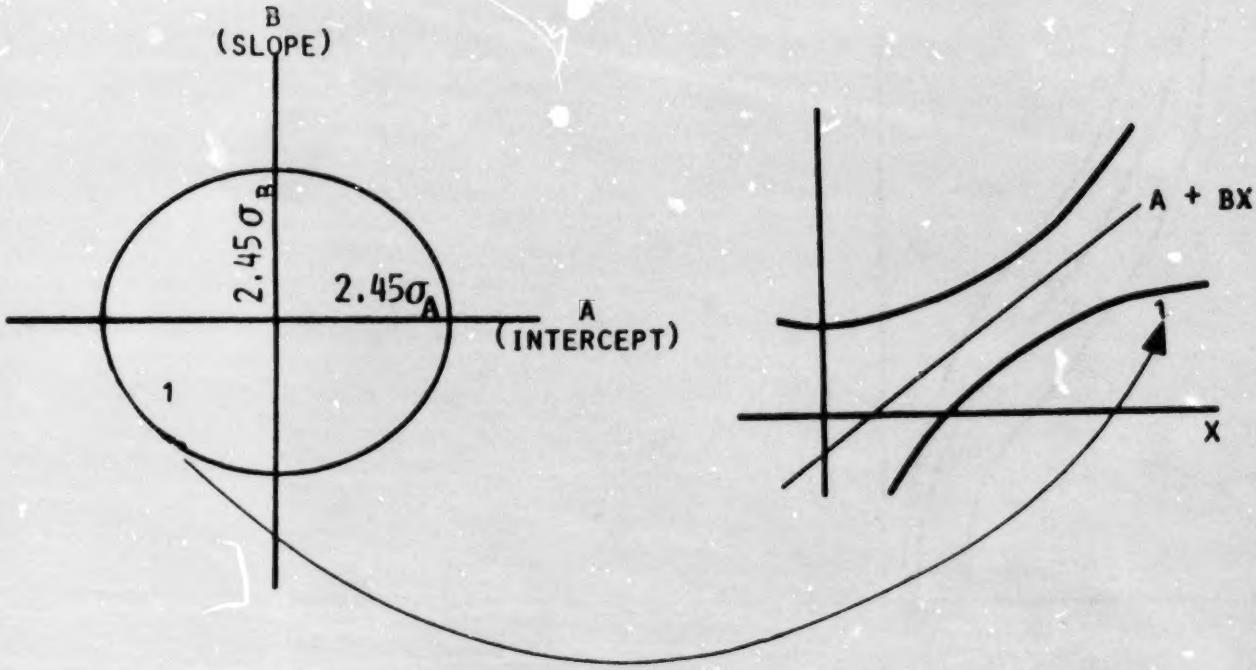
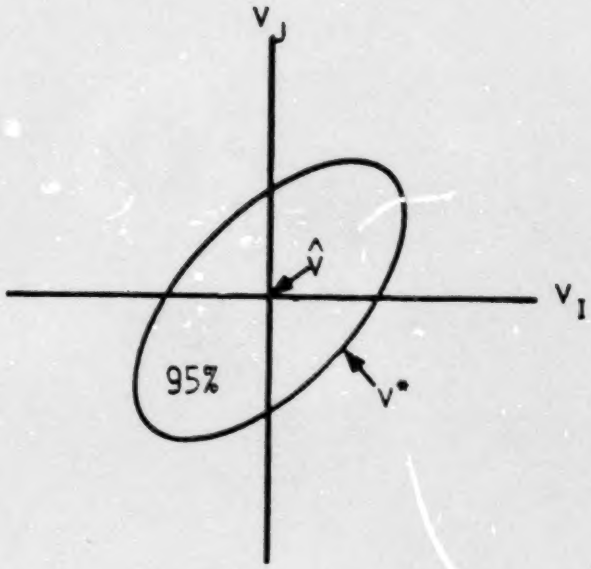


FIGURE 2


v_J
$\hat{v}$
v_I
95%
v^*

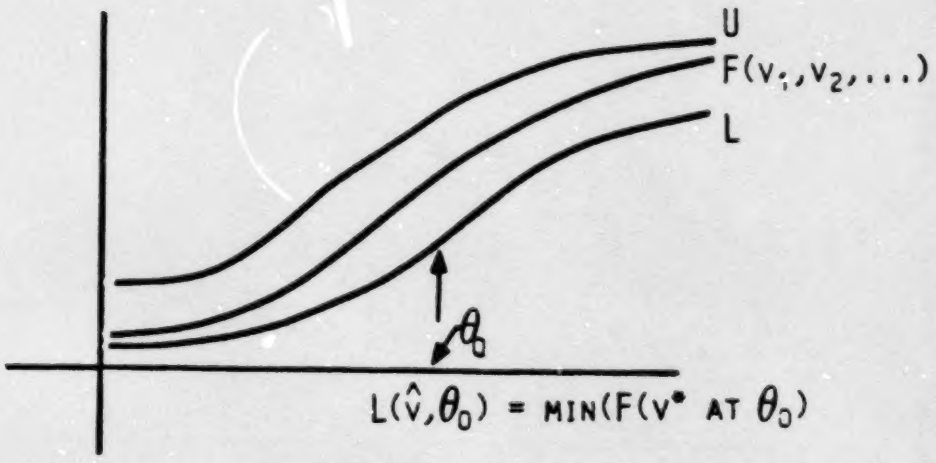

U
$F(v_1, v_2, \ldots)$
L
θ_0
$L(\hat{v}, \theta_0)$ = MIN(F(v^* AT θ_0)

FIGURE 3

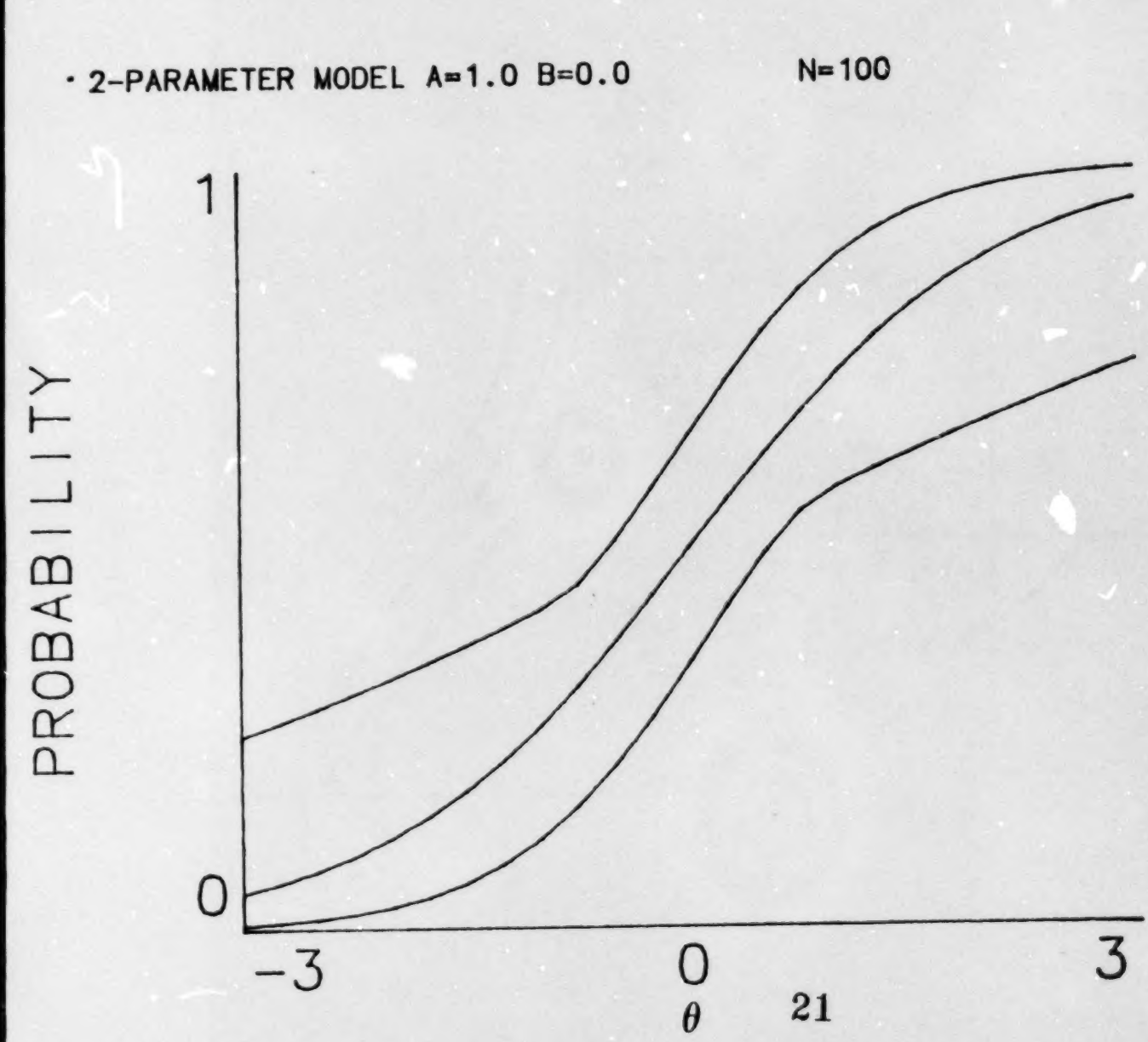


FIGURE 4

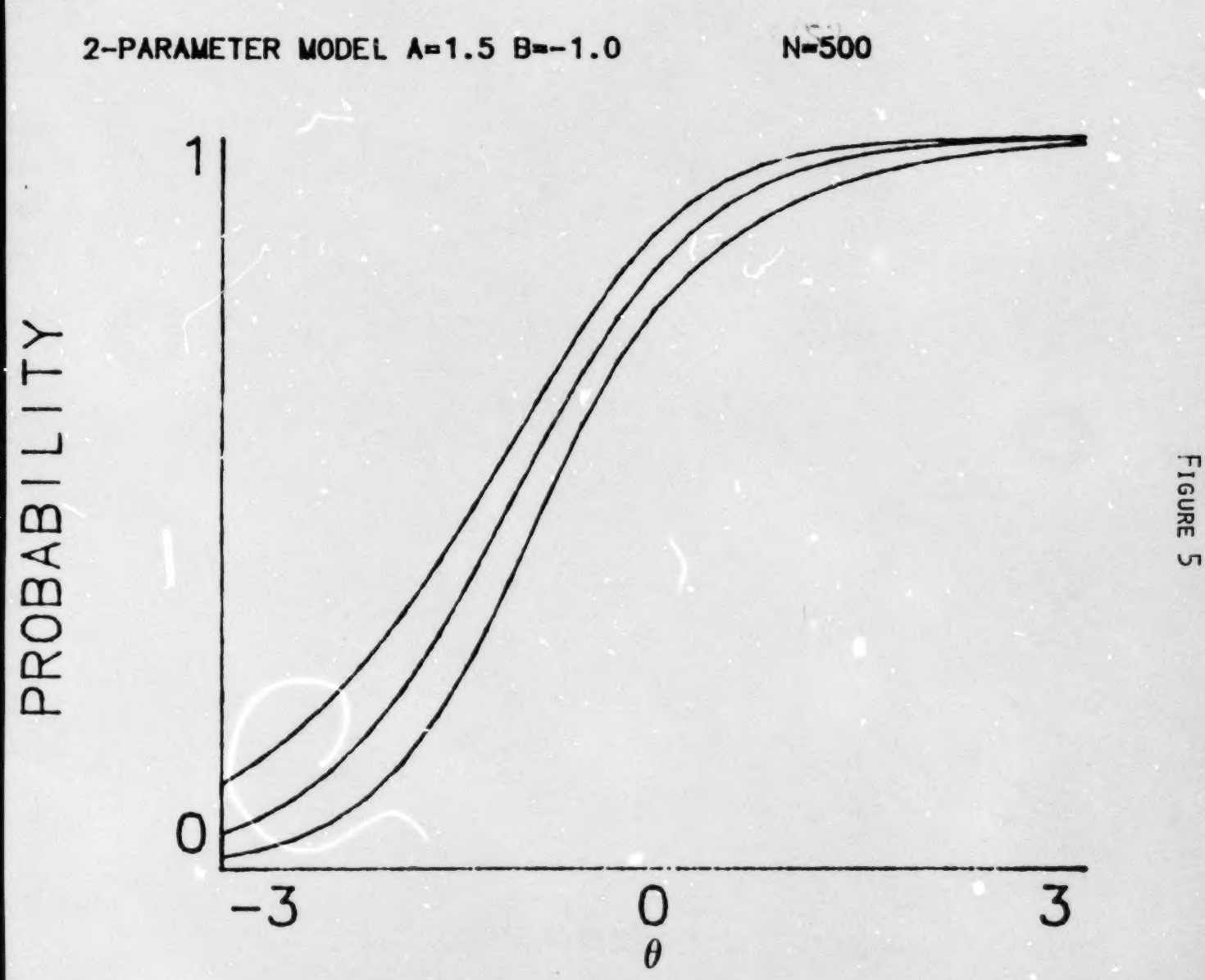


FIGURE 5

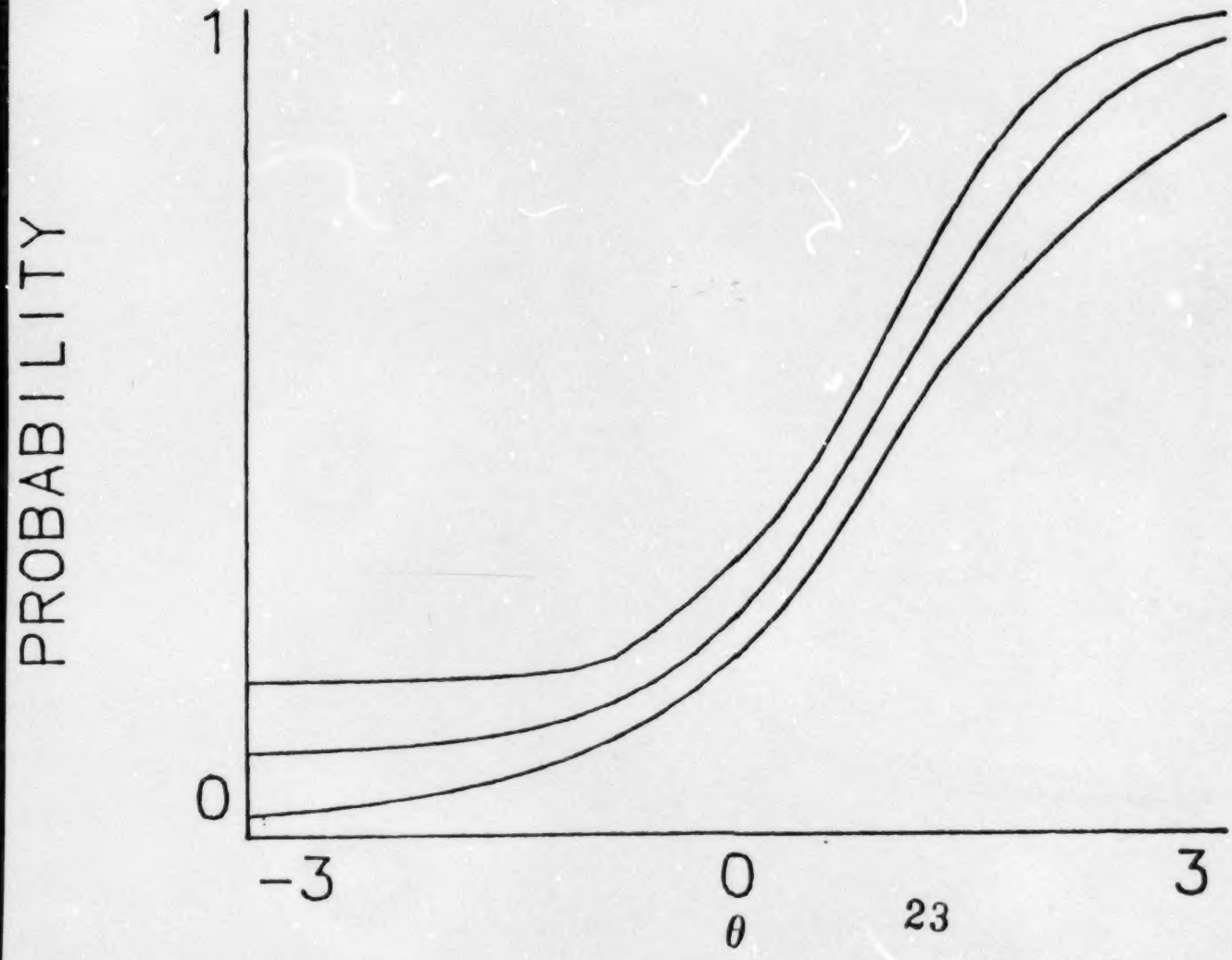


FIGURE 5

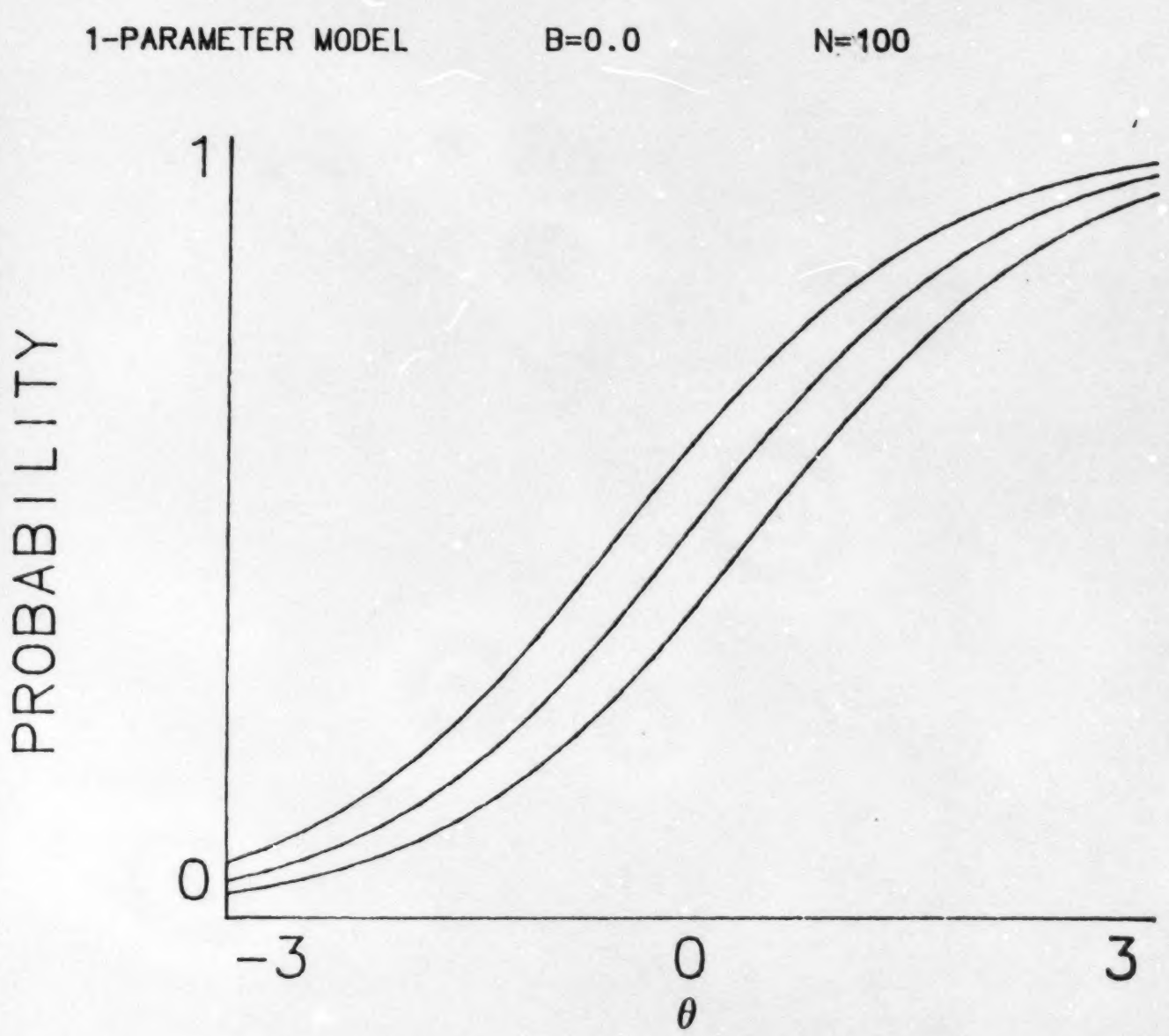


FIGURE 7

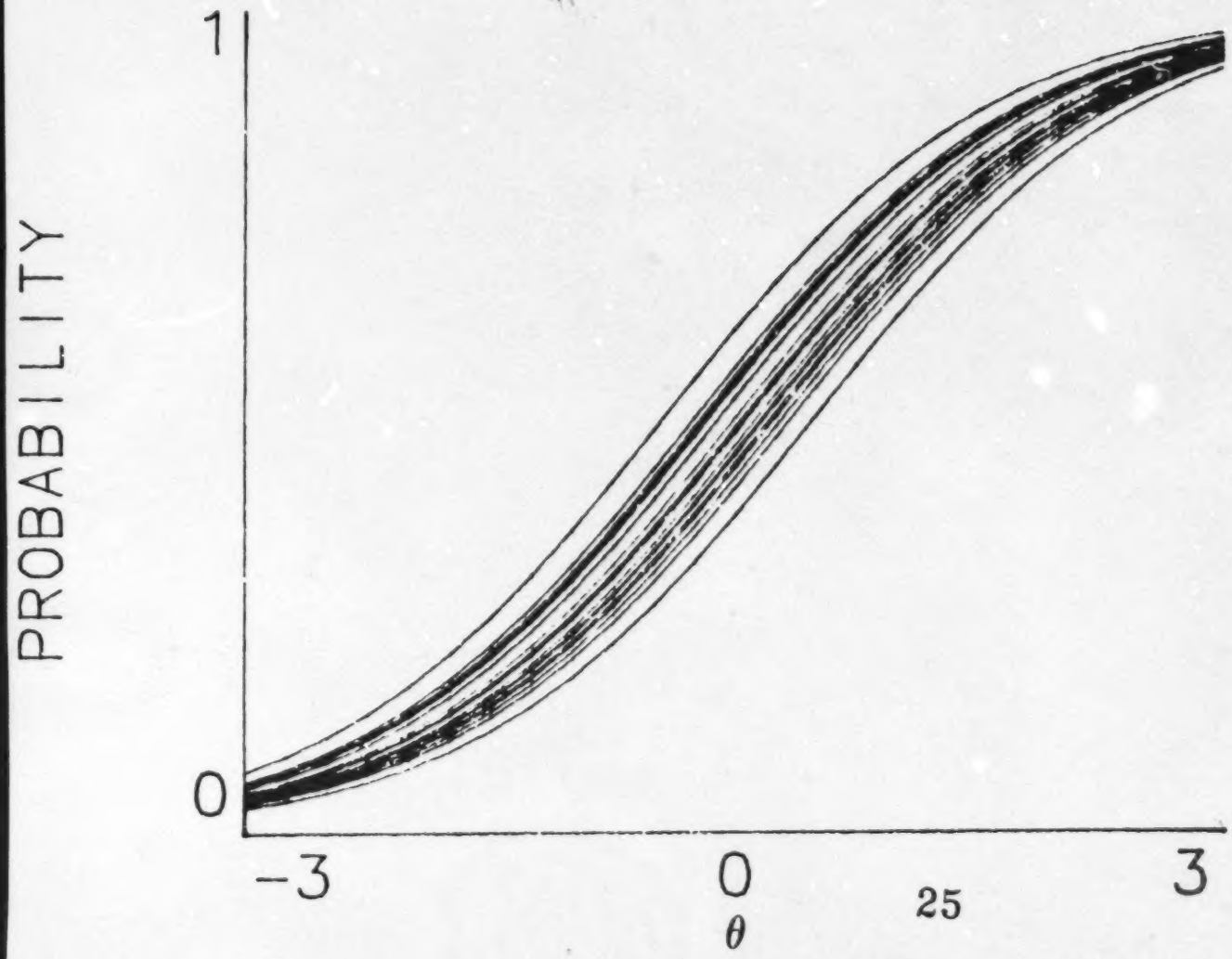


FIGURE 8

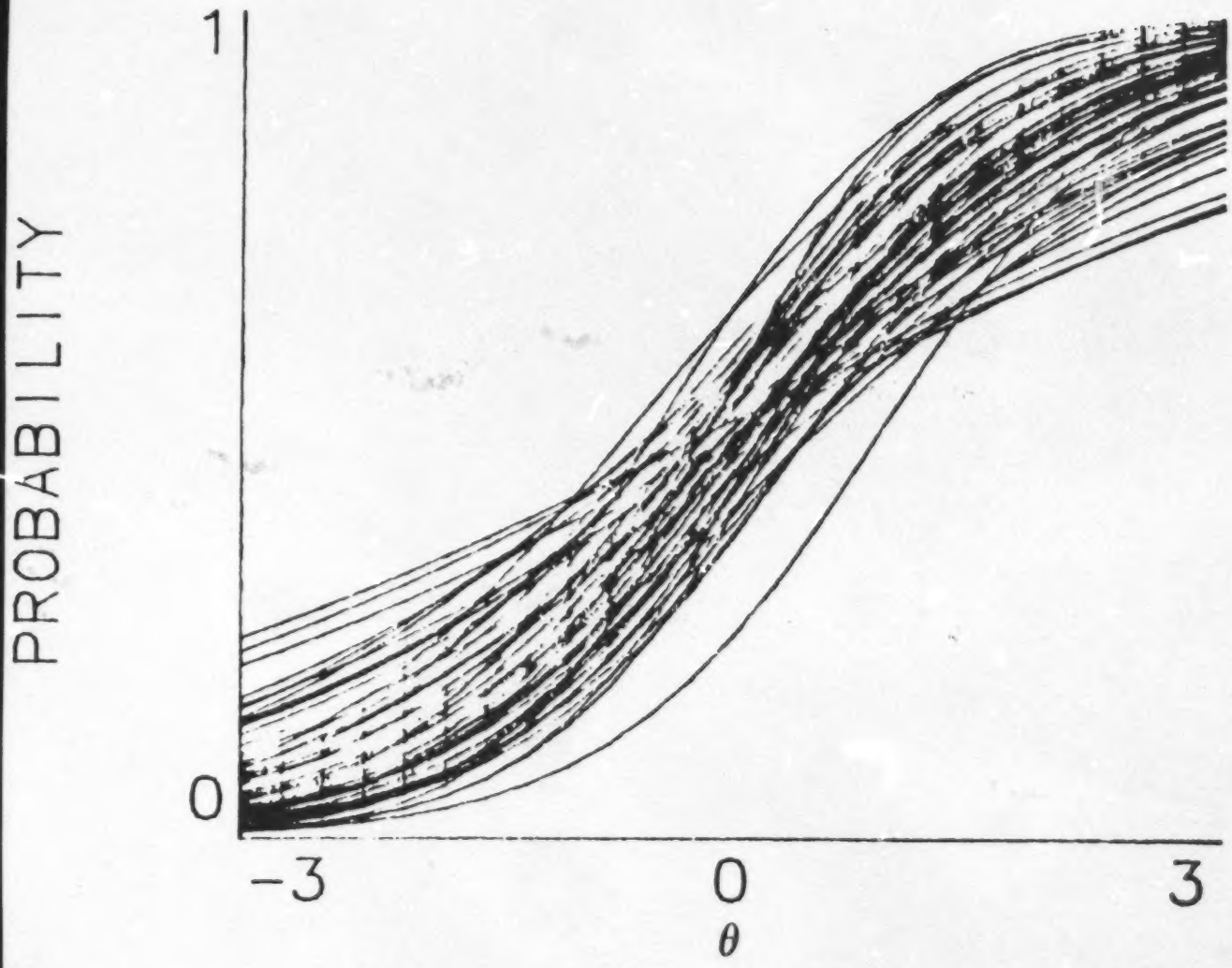


FIGURE 9

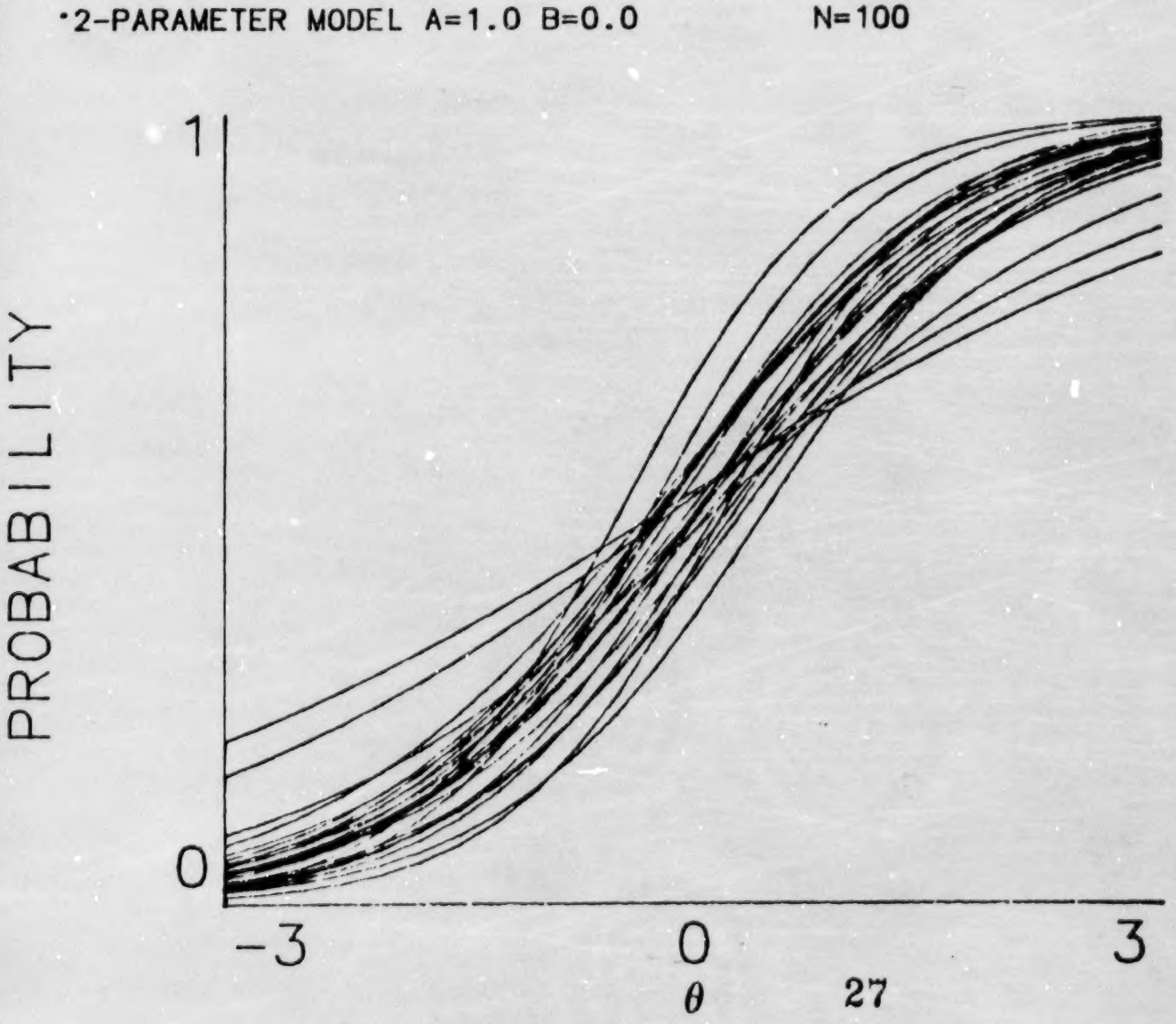


FIGURE 10

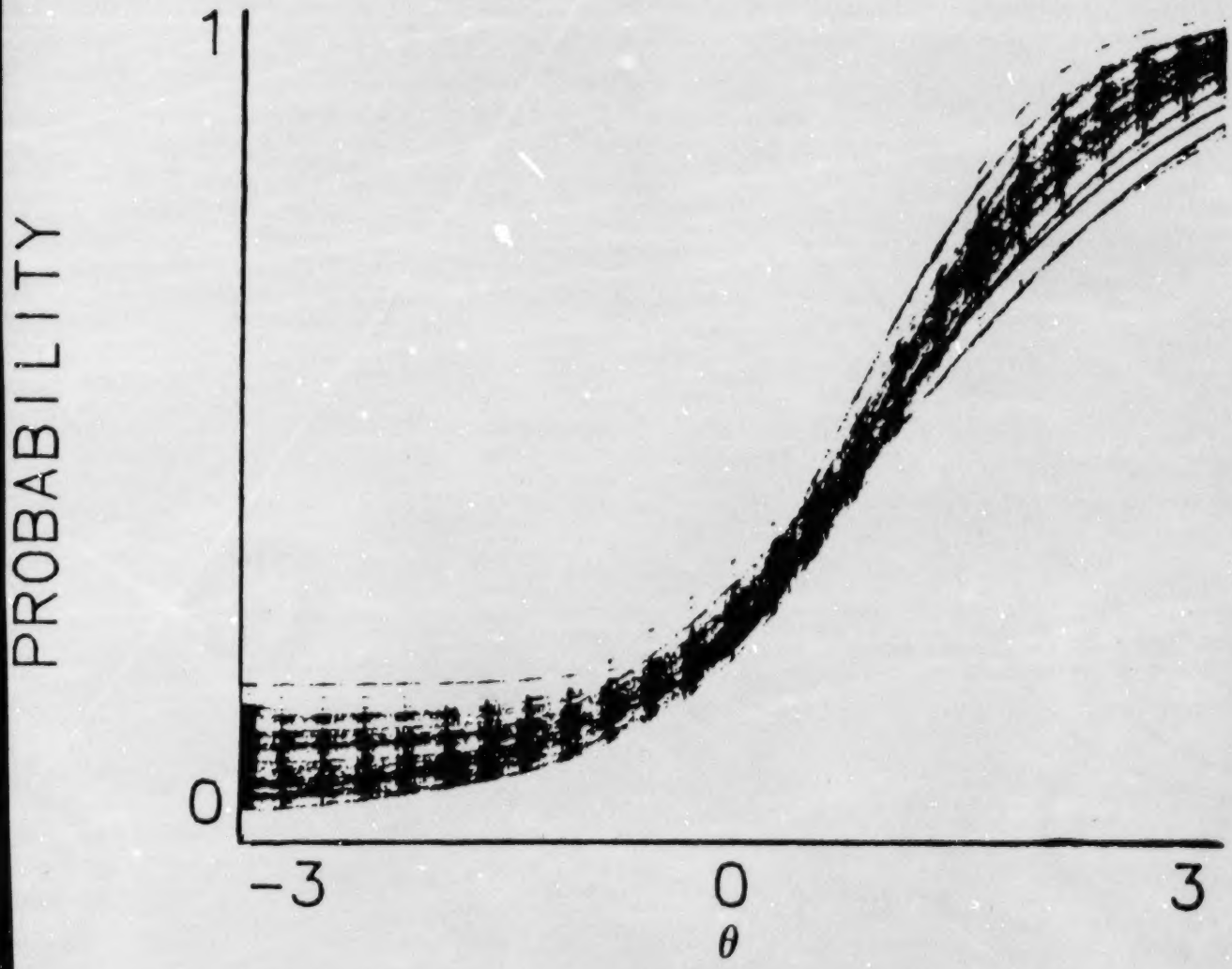


FIGURE 11

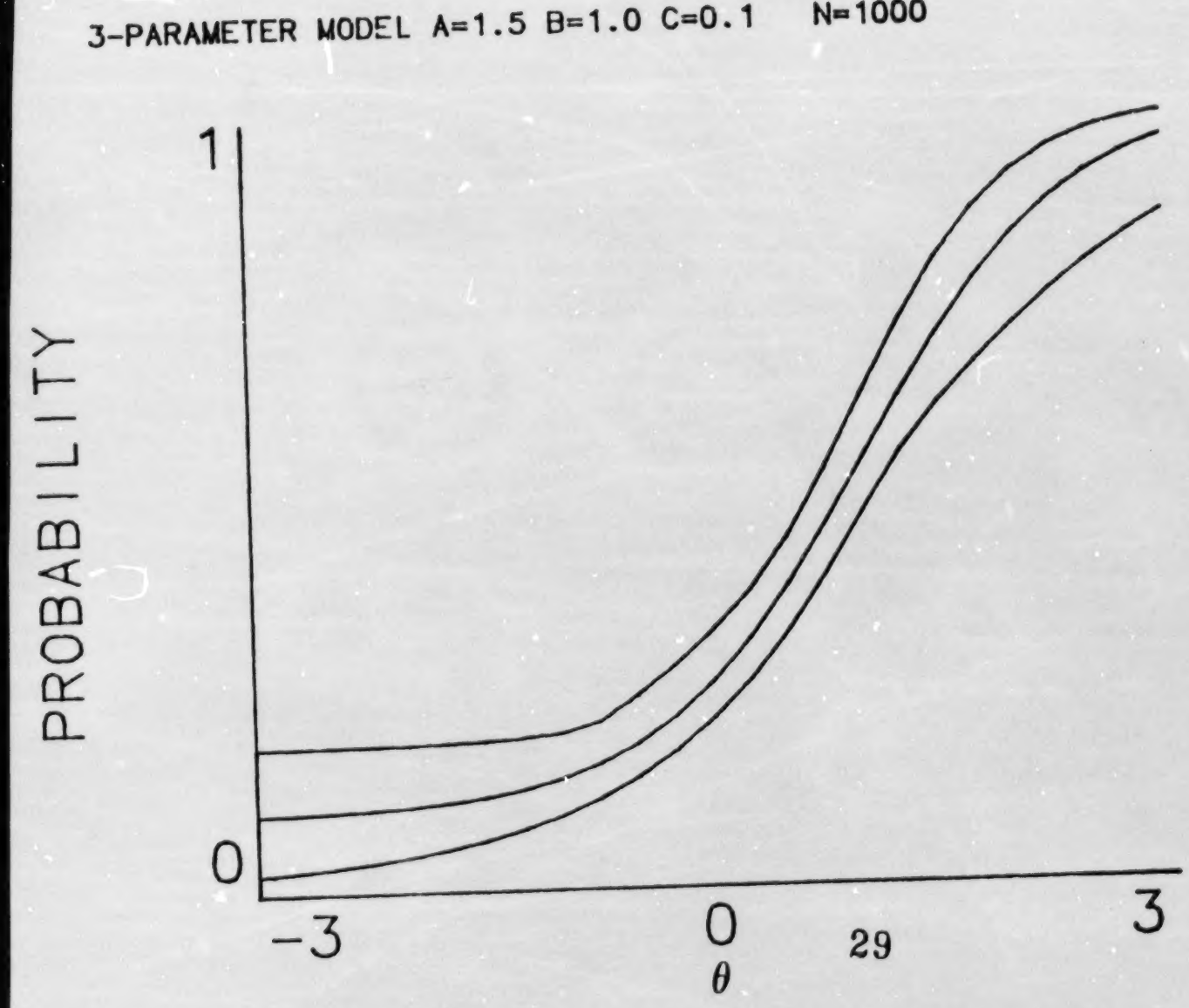


FIGURE 12